

Penerapan Metode *K-Nearest Neighbor* dalam Memprediksi Masa Studi Mahasiswa (Studi Kasus : Mahasiswa STIKOM Dinamika Bangsa)

Jasmir
STIKOM Dinamika
Bangsa
Jambi Indonesia
ijay_jasmir@yahoo.com

Dodo Zaenal Abidin
STIKOM Dinamika
Bangsa
Jambi Indonesia
dikdodoza@gmail.com

Siti Nurmaini
Faculty of Computer
Science Sriwijaya
University, Palembang
Indonesia
sitinurmaini@gmail.com

Reza Firsandaya Malik
Faculty of Computer
Science Sriwijaya
University, Palembang
Indonesia
rezafm@unsri.ac.id

Abstrak- Ledakan pertumbuhan data mahasiswa yang terjadi dalam lembaga perguruan tinggi akan menjadi tumpukan data yang sangat luar biasa. Dengan adanya tumpukan data ini penulis mencoba memanfaatkannya untuk mencari informasi baru. Di STIKOM Dinamika Bangsa tumpukan data yang digunakan merupakan data akademik yaitu data IP (Indeks Prestasi) mahasiswa dari semester satu sampai semester enam. Dengan adanya data ini, penulis melakukan prediksi tentang masa studi mahasiswa. Data dari semester satu sampai semester enam ini akan menjadi dasar perhitungan prediksi yang penulis lakukan dengan tujuan dapat menemukan informasi mahasiswa yang bisa lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu dengan menggunakan salah satu teknik *data mining* yaitu metode klasifikasi *k-Nearest Neighbor* dengan bantuan *Microsoft excel*. Hasil dari penelitian ini adalah nilai prediksi kelulusan mahasiswa dengan berbagai nilai k nya.

Kata Kunci : *Data Mining*, *K-NN*, Prediksi

I. PENDAHULUAN

Salah satu fakta yang terjadi dalam lembaga pendidikan tinggi adalah ledakan pertumbuhan data mahasiswa. Ini terjadi setiap tahun dan setiap penerimaan mahasiswa baru. Data ini meningkat dengan cepat tanpa banyak diketahui manfaatnya kepada manajemen. Pertumbuhan yang pesat dari pertambahan data mahasiswa ini telah menciptakan

beralamat di Jl. Jendral Sudirman Thehok Jambi selatan, yang setiap tahunnya menerima mahasiswa baru rata rata sekitar 500 mahasiswa. Data ini akan bertambah setiap tahunnya dan tentunya akan menjadi tumpukan data.

Dengan adanya tumpukan data di perguruan tinggi yang belum digunakan secara optimal, maka penulis memanfaatkan tumpukan data tersebut untuk mencari informasi baru. Tumpukan data yang digunakan merupakan data akademik di Sekolah Tinggi Ilmu Komputer (STIKOM) Dinamika Bangsa Jambi yaitu data IP (Indeks Prestasi) mahasiswa dari semester satu sampai semester enam. Dengan adanya data ini, penulis berencana melakukan prediksi tentang masa studi mahasiswa. Data dari semester satu sampai semester enam ini akan menjadi dasar perhitungan prediksi yang penulis lakukan dengan harapan dapat menemukan informasi mahasiswa yang bisa lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu dengan menggunakan salah satu teknik *data mining* yaitu metode klasifikasi *k-Nearest Neighbor*.

Data yang digunakan adalah data mahasiswa angkatan 2013 yang baru saja menyelesaikan semester 6, data ini digunakan sebagai data target. Sementara disisi lain juga digunakan data mahasiswa angkatan 2009 sebagai data pembandingan. Data pembandingan ini merupakan data mahasiswa yang sudah lulus.

pada saat ini, tumpukan data tersebut belum dimanfaatkan secara maksimal. Padahal tumpukan data tersebut dapat menjadi sebuah informasi yang bermanfaat dengan menggunakan suatu teknik yaitu teknik *data mining*. Penggunaan teknik *data mining* pada perguruan tinggi dapat berguna mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

STIKOM Dinamika Bangsa Jambi merupakan salah satu perguruan tinggi swasta di kota Jambi yang

2.1. Data Mining

Data mining sering juga disebut *Knowledge Discovery in Database*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [1]

Menurut Turban et al dalam buku Kusriani (2009:3) *data mining* adalah : suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di

dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar”.[2]

Selain beberapa definisi diatas beberapa definisi juga diberikan seperti, Pramudiono dalam buku Kusri (2009:4) mendefinisikan “*Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya”.[3]

2.2. Metode Nearest Neighbor

Metode *Nearest Neighbor* adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru (*testing data*) dengan kasus lama (*training data*), yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. [3]

Jenis metode *Nearest Neighborhood* ada 2, yaitu:

1. 1-NN, yaitu pengklasifikasian dilakukan terhadap 1 *labeled data* terdekat.
2. *k*-NN, yaitu pengklasifikasian dilakukan terhadap *k* *labeled data* terdekat dengan $k > 1$.

Penulis akan menggunakan Metode-*Nearest Neighbor* dalam penelitian ini.

Metode *k-Nearest Neighbor* (*k*-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari label *class* pada *k*-NN. Tujuan dari algoritma *k*-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan *training data*. Algoritma *k*-NN bekerja berdasarkan jarak terpendek dari *query instance* ke *training data* untuk menentukan *k*-NN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *euclidian distance*.

Euclidian Distance sering digunakan untuk menghitung jarak. *Euclidian Distance* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek, di bawah ini merupakan rumus *Euclidian Distance*:

$$\left(\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right)^{1/2} \quad (1)$$

Dimana,

X_{ik} = nilai X pada *training data*

X_{jk} = nilai X pada *testing data*

m = batas jumlah banyaknya data

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah *training data* dan *testing data*.

Dalam algoritma ini, nilai k yang terbaik itu tergantung pada jumlah data. Ukuran nilai k yang besar belum tentu menjadi nilai k yang terbaik begitupun juga sebaliknya. Langkah-langkah untuk menghitung algoritma *k*-NN:

1. Menentukan nilai k .
 2. Menghitung kuadrat jarak *euclid* (*query instance*) masing-masing objek terhadap *training data* yang diberikan.
 3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclid* terkecil.
 4. Mengumpulkan label *class* Y (klasifikasi *Nearest Neighborhood*).
 5. Dengan menggunakan kategori *Nearest Neighborhood* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung.
- ### III. METODOLOGI PENELITIAN

Kerangka kerja dalam melakukan penelitian (metodologi Penelitian) ini adalah sebagai berikut :



Gambar 1 Kerangka Kerja

1. Perumusan Masalah

Masalah yang dirumuskan dalam penelitian ini adalah bagaimana penerapan metode *K-Nearest Neighbor* dalam memprediksi masa studi mahasiswa STIKOM Dinamika Bangsa Jambi.

2. Penentuan Tujuan

Tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi metode *K-Nearest Neighbor* dalam memprediksi masa studi mahasiswa

3. Mempelajari Literatur

Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, literatur-literatur bersumber dari buku-buku perpustakaan STIKOM Dinamika Bangsa Jambi dan jaringan internet. Literatur-literatur yang digunakan nanti dilampirkan dalam daftar pustaka.

4. Pengumpulan Data dan Informasi

Dalam pengumpulan data, penulis mendapatkan data melalui bagian BAAK STIKOM Dinamika Bangsa Jambi..

5. Proses Data Mining

Data Mining adalah proses pengekstrasian *knowledge* yang tersimpan dalam database bervolume besar. Untuk mendapatkan *knowledge* dalam database digunakanlah metode *K-Nearest Neighbord*.

6. Pengujian

Pada tahap ini dilakukan pengujian dari hasil yang didapat dari tahap sebelumnya sebagai pedoman untuk membantu mahasiswa yang dalam merencanakan kelulusannya secara tepat waktu.

IV. PEMBAHASAN

Dalam penelitian ini akan dicari nilai *euclidean* dari hubungan tingkat kelulusan dengan data induk mahasiswa, yaitu dari nilai IP semester satu sampai semester enam dan akan diuji menggunakan nilai IP di dua semester, empat semester dan enam semester sehingga hasil dari proses tersebut dapat dilihat dengan menggunakan IP disemester berapa dan nilai *k* yang terbaik yang menghasilkan tingkat keberhasilannya tinggi.

4.1. Sumber Data

Data yang digunakan dalam penelitian ini menggunakan data akademik. Data akademik ini dibuat menjadi dua data yaitu *training data* dan *testing data*. *Training data* dan *testing data* merupakan data mahasiswa yang didata ketika mahasiswa tersebut telah menjadi mahasiswa di

Program Studi Teknik Informatika dan yang telah lulus.

Data akademik mahasiswa yang diambil adalah data mahasiswa angkatan 2009. Hal ini didasarkan pada kebutuhan data yang akan dihubungkan dengan *training data*, dengan asumsi bahwa mahasiswa angkatan 2009 akan lulus dari rentang waktu tahun 2013-2018. Sedangkan data kelulusan di *testing data* rentang waktunya dari tahun 2013. Kedua data tersebut diperoleh dari BAAK STIKOM Dinamika Bangsa Jambi. Data yang diambil hanya dari data mahasiswa Program Studi Teknik Informatika.

4.2 Penggunaan Algoritma *k*-NN

Algoritma *k*-NN merupakan algoritma untuk mencari jarak terdekat antara *training data* dengan *testing data*. Untuk menghitung jarak terdekat atau jauhnya tetangga dapat menggunakan metode *euclidean distance*.

Contoh proses *data mining* menggunakan algoritma *k*-NN yaitu sebagai berikut:

Terdapat beberapa data yang berasal dari IP mahasiswa Teknik Informatika yang telah lulus sebagai *training data* (Tabel 4.3) untuk diklasifikasikan dengan *testing data* menggunakan enam atribut yaitu IP dari semester satu sampai semester enam (Tabel 4.4) sehingga dapat menentukan masa studi mahasiswa apakah mahasiswa tersebut lulus tepat waktu atau tidak.

Tabel 4.1. Tabel Data Mahasiswa angkatan 2009

No.	NIM	Nama Mahasiswa	sem1	sem2	sem3	sem4	sem5	sem6
1.	8020090001	Yeyen Apriyanti	3,4	3,63	3,66	3,25	3,90	3,70
2.	8020090002	Nita Andila	3,77	3,81	3,95	3,83	3,81	4,00
3.	8020090003	Fajar	3,45	3,72	3,83	3,66	3,81	3,87
4.	8020090004	Hendri	2,72	3,68	3,41	3,08	3,45	3,54
...
220.	8020090290	Andri Setiawan	2,18	2,54	1,77	2,87	2,95	2,86
221.	8020090291	Apin Sutikno	2,45	2,90	2,95	2,50	3,22	3,09
222.	8020090294	Muhammad Amin	0,36	2,86	3,13	3,36	3,41	3,00
223.	8020090295	Robin Sitepu	0,36	3,54	3,33	3,41	3,45	3,50

Tabel 4.2 Tabel Atribut Data Mahasiswa angkatan 2009

Atribut	Keterangan
NIM	Nomor Induk Mahasiswa
Nama	Nama Mahasiswa
Sem	Digunakan untuk proses <i>mining</i> guna mengetahui hubungan antara tingkat kelulusan dengan IP dari semester 1-6.

Tabel 4.3. Tabel Training Data Lengkap

No.	NIM	sem1a	sem2a	sem3a	sem4a	sem5a	sem6a	Y=Class
1.	8020090001	3,4	3,63	3,66	3,25	3,90	3,70	Ya
2.	8020090002	3,77	3,81	3,95	3,83	3,81	4,00	Ya
3.	8020090003	3,45	3,72	3,83	3,66	3,81	3,87	Ya
4.	8020090004	2,72	3,68	3,41	3,08	3,45	3,54	Ya
...
221.	8020090291	2,45	2,90	2,95	2,50	3,22	3,09	Tidak
222.	8020090294	0,36	2,86	3,13	3,36	3,41	3,00	Tidak
223.	8020090295	0,36	3,54	3,33	3,41	3,45	3,50	Tidak

Keterangan:

Klasifikasi= keterangan yang menyatakan mahasiswa yang telah lulus kurang atau tepat 5 tahun (ya) dan lebih dari 5 tahun (tidak).

Tabel 4.4. Tabel Testing Data Lengkap

No.	NIM	Nama Mahasiswa	Sem1b	Sem2b	Sem3b	Sem4b	Sem5b	Sem6b	Y=Class
1.	8020130001	Calvinia Flora	3,62	3,71	3,64	3,91	3,93	3,76	?
2.	8020130002	Stella Octavius	3,75	3,80	3,79	3,76	3,93	3,86	?
3.	8020130003	Ani Safitri	3,75	3,88	3,87	3,54	4,00	3,86	?
4.	8020130004	Ivan Viriya W	3,92	3,90	3,91	3,47	3,77	3,75	?
...
219.	8020130278	Gunawan Tri	3,67	3,30	3,43	3,73	3,54	3,72	?
220.	8020130279	Hendri Manalu	2,72	2,92	3,11	3,14	3,50	3,28	?
221.	8020130282	Wahyuzi A	2,80	2,97	3,26	3,21	3,47	3,06	?

Testing data merupakan sekumpulan data IP mahasiswa yang akan diklasifikasikan dengan *training data*, apakah lulus ≤ 5 tahun atau lulus > 5 tahun.

Setelah ada *testing data* dan *training data*, lalu menentukan nilai k -nya, contoh untuk kasus di atas nilai k yang digunakan adalah $k=5$.

Adapun langkah-langkah menjawab permasalahan di atas, yaitu:

1. Nilai k yang digunakan adalah 5.
2. Menghitung kuadrat jarak *euclid* (*query instance*) masing-masing objek terhadap sampel data atau *training data* yang diberikan dengan menggunakan rumus (1)

Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *Euclid* terkecil.

Pada table 4.7, yang dimaksud dengan jarak terkecil adalah mengurutkan hasil *query instance* dari yang terkecil hingga ke besar. Jika urutan pada *query instance* termasuk nilai $k = 5$, maka yang diambil 5 jarak yang terpendek yang termasuk nilai k (Ya).

Selanjutnya pada proses berikutnya hasil dari pengelompokan “apakah termasuk *Nearest Neighbor* (k)” akan dikelompokkan dengan “Y=Klasifikasi” untuk menentukan data tersebut apakah termasuk klasifikasi lulus ≤ 5 tahun atau lulus > 5 tahun.

Dengan menggunakan kategori *Nearest Neighborhood* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung. Pada urutan jarak yang terdekat dari satu sampel lima (nilai $k=5$), maka diketahui ada 2 yang lulus ≤ 5 tahun dan 3 yang lulus > 5 tahun pada Tabel 3.10. Sehingga *testing data* tersebut termasuk lulus > 5 tahun.

4.3 PENGUJIAN DATA

Pengujian terdiri dari 2 proses yaitu

1. Pengujian 1 *database* mahasiswa (keseluruhan) dengan menggunakan *data training* yang berjumlah 223 data.

2. Pengujian 1 *database* mahasiswa (keseluruhan) dengan menggunakan *data testing* yang berjumlah 221 data.

Masing-masing proses pengujian tersebut menggunakan Indeks Prestasi (IP) dua semester (semester 1 dan 2), empat semester (semester 1-4) dan enam semester (semester 1-6) dengan menggunakan nilai k yang berbeda. Nilai k yang digunakan yaitu kelipatan 5 untuk *training data* berjumlah 221 data dan nilai k yang digunakan untuk *training data* berjumlah 221 adalah kelipatan 5 juga. Hasil prediksi 1 *database* akan dibandingkan dengan data asli dan dicari kecocokannya secara otomatis. Pengujian ini juga berguna untuk mengetahui apakah nilai k yang digunakan adalah nilai k yang terbaik dengan hasil tingkat keberhasilannya tinggi atau tidak untuk memprediksi kelulusan mahasiswa. Untuk mengetahui tingkat keberhasilan pada sistem ini maka digunakan rumus di bawah ini :

Tingkat keberhasilan

$\frac{\text{Hasil pengujian bernilai benar (kecocokan)}}{\text{Banyaknya data sampel}}$

----- x
100%

$\frac{\text{Hasil pengujian bernilai benar (kecocokan)}}{\text{Banyaknya data sampel}}$

1. Nilai k Menggunakan Nilai IP Dua Semester

Pengujian ini dilakukan dengan cara menggunakan nilai IP disemester satu dan dua dengan tujuan dapat mengetahui nilai k yang terbaik dan persentase tingkat keberhasilannya.

Pada Tabel 4.10 menunjukkan beberapa hasil pengujian sistem aplikasi *data mining* dengan mengubah-ubah nilai k . Untuk melihat hasil pengujian secara lengkap dapat dilihat di lampiran hasil pengujian.

Tabel 4.10 Hasil Pengujian Nilai k Menggunakan Nilai IP Dua Semester

Hasil Prediksi		
K = 5	K = 10	K = 15

51%	69%	58%
-----	-----	-----

Dari data Tabel 3.10 terlihat bahwa nilai k yang terbaik untuk memprediksi masa studi mahasiswa menggunakan Indeks Prestasi (IP) semester satu dan dua dengan menggunakan *training data* yang berjumlah 30 data adalah nilai $k=10$ dengan tingkat keberhasilan 69.00%.

2. Pengujian Nilai k dalam Empat Semester

Pengujian ini dilakukan dengan cara menggunakan nilai IP disemester satu sampai empat dengan tujuan dapat mengetahui nilai k yang terbaik dan persentase tingkat keberhasilannya. Pada Tabel 3.11 di bawah ini menunjukan beberapa hasil pengujian sistem aplikasi *data mining* dengan mengubah-ubah nilai k . Untuk melihat hasil pengujian secara lengkap dapat dilihat di lampiran hasil pengujian.

Tabel 3.11 Hasil Pengujian Nilai k Menggunakan Nilai IP Empat Semester

Hasil Prediksi		
K = 5	K = 10	K = 15
59%	63%	74%

Dari data Tabel 3.11 diatas terlihat bahwa nilai k yang terbaik untuk memprediksi masa studi mahasiswa menggunakan Indeks Prestasi (IP) semester satu samapai empat dengan menggunakan *data training* yang berjumlah 223 data adalah nilai $k=15$ dengan tingkat keberhasilan 74.00%.

3. Pengujian Nilai K Dalam Enam Semester

Pengujian ini dilakukan dengan cara menggunakan nilai IP dari semester satu sampai enam dengan tujuan dapat mengetahui nilai k yang terbaik dan tingkat keberhasilan jika mengklasifikasi masa studi mahasiswa. Pada Tabel 3.12 di bawah ini menunjukan

beberapa hasil pengujian sistem aplikasi *data mining* dengan mengubah-ubah nilai k . Untuk melihat hasil pengujian secara lengkap dapat dilihat di lampiran hasil pengujian.

Tabel 5.12 Hasil Pengujian Nilai k Menggunakan Nilai IP Enam Semester

Hasil Prediksi		
K = 5	K = 10	K = 15
76%	83%	82%

V. KESIMPULAN

Berdasarkan pembahasan dari bab-bab sebelumnya, maka penelitian diambil beberapa simpulan:

1. ProsesProses Prediksi masa studi mahasiswa menggunakan algoritma *k-Nearest Neighbor* yang akan diterapkan di Jurusan Teknik Informatika.
2. Untuk menggunakan *training data* yang berjumlah 223 data dengan menguji *testing data* berjumlah 221 data, maka didapatkan nilai k yang terbaik untuk memprediksi masa studi mahasiswa yaitu sebagai berikut:
 - a. Untuk dua semester yaitu nilai k yang terbaik untuk digunakan memprediksi studi mahasiswa adalah nilai $k = 10$ dengan tingkat keberhasilan 69%.
 - b. Untuk empat semester yaitu nilai k yang terbaik untuk digunakan memprediksi studi mahasiswa adalah nilai $k = 15$ dengan tingkat keberhasilan 74%.

Untuk enam semester yaitu nilai $k = 10$ merupakan nilai k yang terbaik untuk digunakan memprediksi masa studi mahasiswa dengan tingkat keberhasilan 83%.

Tabel 4.5. Tabel Testing Data uji coba 1

7.	8020130008	Elise Aprilia	4,00	3,85	4,00	3,95	4,00	3,84	?
----	------------	---------------	------	------	------	------	------	------	---

Tabel 4.6 Square Instance to Query Distance

No.	NIM	sem1	sem2	sem3	sem4	sem5	sem6	SlQD
1.	8020090001	3,4	3,63	3,66	3,25	3,90	3,70	1,021567
2.	8020090002	3,77	3,81	3,95	3,83	3,81	4,00	0,364829
3.	8020090003	3,45	3,72	3,83	3,66	3,81	3,87	0,685128
4.	8020090004	2,72	3,68	3,41	3,08	3,45	3,54	1,778988
...
221.	8020090291	2,45	2,90	2,95	2,50	3,22	3,09	2,771444
222.	8020090294	0,36	2,86	3,13	3,36	3,41	3,00	4,048259
223.	8020090295	0,36	3,54	3,33	3,41	3,45	3,50	3,808451

Tabel 4.7 Urutan Objek ke dalam Kelompok ke Jarak Euclid Terkecil

No	NIM	sem1	sem2	sem3	sem4	sem5	sem6	Rata	SlQD	UJT	AkNN
1.	8020090001	3,4	3,63	3,66	3,25	3,90	3,70	3,59	1,022	25	Tidak

2.	8020090002	3,77	3,81	3,95	3,83	3,81	4,00	3,86	0,365	3	Ya
3.	8020090003	3,45	3,72	3,83	3,66	3,81	3,87	3,72	0,685	12	Tidak
4.	8020090004	2,72	3,68	3,41	3,08	3,45	3,54	3,31	1,779	67	Tidak
...
221.	8020090291	2,45	2,90	2,95	2,50	3,22	3,09	2,85	2,771	110	Tidak
222.	8020090294	0,36	2,86	3,13	3,36	3,41	3,00	2,69	4,048	171	Tidak
223.	8020090295	0,36	3,54	3,33	3,41	3,45	3,50	2,93	3,808	147	Tidak

Keterangan

SitQD : Square Instance to Query Distance

UJT : Urutan Jarak Terkecil

AkNN : Apakah Termasuk Nearest Neighbor

Tabel 4.8 Tabel Label Class Y

No	NIM	sem1	sem2	sem3	sem4	sem5	sem6	Rata	SitQD	UJT	AkNN	Y=Class
1.	8020090001	3,4	3,63	3,66	3,25	3,90	3,70	3,59	1,022	25	Tidak	Ya
2.	8020090002	3,77	3,81	3,95	3,83	3,81	4,00	3,86	0,365	3	Ya	Ya
3.	8020090003	3,45	3,72	3,83	3,66	3,81	3,87	3,72	0,685	12	Tidak	Ya
4.	8020090004	2,72	3,68	3,41	3,08	3,45	3,54	3,31	1,779	67	Tidak	Ya
...
221.	8020090291	2,45	2,90	2,95	2,50	3,22	3,09	2,85	2,771	110	Tidak	Tidak
222.	8020090294	0,36	2,86	3,13	3,36	3,41	3,00	2,69	4,048	171	Tidak	Tidak
223.	8020090295	0,36	3,54	3,33	3,41	3,45	3,50	2,93	3,808	147	Tidak	Tidak

Tabel 4.9 Tabel hasil akhir klasifikasi

No.	NIM	Sem1	Sem2	Sem3	Sem4	Sem5	Sem6	Klasifikasi
1.	8020130001	3,62	3,71	3,64	3,91	3,93	3,76	Ya
2.	8020130002	3,75	3,80	3,79	3,76	3,93	3,86	Ya
3.	8020130003	3,75	3,88	3,87	3,54	4,00	3,86	Ya
4.	8020130004	3,92	3,90	3,91	3,47	3,77	3,75	Ya
...
219.	8020130278	3,67	3,30	3,43	3,73	3,54	3,72	Ya
220.	8020130279	2,72	2,92	3,11	3,14	3,50	3,28	Ya
221.	8020130282	2,80	2,97	3,26	3,21	3,47	3,06	Ya

REFERENSI

- [1]. Budi Santosa, 2007 Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Teori dan Aplikasi . Yogyakarta, Graha Ilmu
- [2]. Turban, Efraim, dkk. 2005. *Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas)* Edisi 7. Yogyakarta : Andi Offset.
- [3]. Kusriani, Emha Taufiq Luthfi. 2009 *Algoritma Data Mining* .Yogyakarta : andi Offset
- [4]. Han, J. et al. 2006 *Data Mining: Concepts and Techniques*, 2nd Edition, Morgan Kaufmann Publisher,
- [5]. Muhammad Syukri Mustafa, dkk. 2014 "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar)" *Citec Journal*, Vol. 1, No. 4, pp 270-281
- [6]. Mutiara Ayu Banjarsari, dkk, 2015 " Penerapan K-Optimal Pada Algoritma Knn untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4" *Kumpulan jurnal Ilmu Komputer (KLIK)* Volume 02, No.02: pp 50-64
- [7]. Sajadin Sembiring, dkk, 2011, *Prediction Of Student Academic Performance By An Application Of Data Mining Techniques*, *International Conference on Management and Artificial Intelligence IPEDR vol.6 (2011)* IACSIT Press, Bali, Indonesia
- [8]. Surjeet Kumar Yadav, 2012, *Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification*, *World of Computer Science and Information Technology Journal (WCSIT)* ISSN: 2221-0741 Vol. 2, No. 2, pp 51-56